

# Integrated Transcriptome-wide Profiling and Protein Structure Analysis of Pathogenic Genes in Venous Thromboembolism

Yan Che\*, Jing Ding, Yan Zhang

Department of Reproduction Regulation, Fudan University, Shanghai, China

## Research Article

**Received:** 21-Jun-2022, Manuscript No. JOB-22-67344; **Editor assigned:** 24-Jun-2022, PreQC No. JOB-22-67344 (PQ); **Reviewed:** 11-Jul-2022, QC No. JOB-22-67344; **Revised:** 22-Jul-2022, Manuscript No. JOB-22-67344 (R); **Published:** 29-Jul-2022, DOI: 10.4172/2322-0066.10.6.002.

### \*For Correspondence

Yan Che, Department of Reproduction Regulation, Fudan University, Shanghai, China

**E-mail:** cheyan2004@163.com

**Keywords:** Venous Thromboembolism; Transcriptome-Wide Profiling; Whole Exome Sequencing; Protein Structure; HNMT

**Keywords:** VTE: Venous Thromboembolism; DEGs: Differentially Expressed Genes; GEO: Gene Expression Omnibus; DEGs: Differentially Expressed Genes; WES: Whole-Exome Sequencing; KEGG: Kyoto Encyclopedia of Genes and Genomes; HNMT: Histamine N-Methyltransferase; PCA: Principal Component Analysis; BPs: Biological Processes; MFs: Molecular Functions; CCs: Cellular Components; SNPs: Single-Nucleotide Polymorphisms; INDELS: Insertion-Deletions.

### ABSTRACT

**Back ground:** Genetic factors are considered to determine the balance of the coagulation and anticoagulation processes, yet the genetic variants related to Venous Thromboembolism (VTE) remain unclear. This study aimed to investigate the potential molecular mechanisms and pathogenic mutations associated with VTE by determining VTE-related Differentially Expressed Genes (DEGs) by transcriptome-wide profiling and assaying protein structure in VTE.

**Methods:** Two gene expression datasets, *GSE48000* and *GSE19151*, were accessed from the Gene Expression Omnibus (GEO) database to obtain gene expression data associated with VTE. We identified the DEGs between VTE patients and healthy people using R and performed functional enrichment analysis, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. Then, Whole-Exome Sequencing (WES) was performed for 25 VTE patients and 17 normal cases, and the structural locations of pathogenic missense mutations were identified using pyMOL. Finally, DGIdb database was used to select candidate drugs for the treatment of VTE.

**Results:** A total of 232 DEGs were identified from the GEO database. The significant function of these DEGs was mostly involved in RNA catabolic process and ribosome pathway. Notably, the results of WES for DEGs and protein structure analysis showed that Histamine N-Methyltransferase (HNMT) (chr2: 138759649 C>T, rs11558538) may be a main predisposing factor for VTE. In addition, Amodiaquine, Harmaline, Aspirin, Metoprine, Dabigatran, and Diphenhydramine were screened for VTE therapy.

**Conclusion:** The results showed that HNMT (chr2: 138759649 C>T, rs11558538) may be potential target for the diagnosis and treatment of VTE.

## INTRODUCTION

Venous Thromboembolism (VTE) is the third most common cardiovascular disease worldwide, which manifests as deep-vein thrombosis, pulmonary embolism, or both <sup>[1,2]</sup>. Various epidemiological studies have demonstrated that the incidence of VTE is characterized by a remarkable number of genetic and environmental factors. In early epidemiological studies, the highest incidence of VTE was in Africa, followed by Caucasians and was lower for Asian <sup>[3]</sup>. With increased awareness of the diagnosis and management of VTE, the incidence in Asian has increased in recent years <sup>[4]</sup>.

While VTE is classified as a complex, multifactorial and polygenic disease, common mechanisms driven VTA have been confirmed, such as gene-gene and gene-environment interactions <sup>[5]</sup>. Stasis, vessel damage, and a hypercoagulable state are three widely accepted mechanisms related to the occurrence of VTE <sup>[6]</sup>. Genetic epidemiological studies have revealed that genetic conditions are significant risk factors for VTE, accounting for up to 50% of all VTE patients including anticoagulant protein deficiency trapping (deficiency of protein C, protein S, and antithrombin), Factor V Leiden mutation (FVL) (c.1601G>A, p.R534Q), prothrombin G20210A mutation (FII G20210A), hyperhomocysteinemia, elevated coagulation factors VIII, IX, X, histidine rich glycoprotein, and ABO blood group <sup>[7,8]</sup>. However, only a few genetic factors have been considered and the distribution of FVL and FII G20210A

mutation depends on ethnic group, race, or geographical region, suggesting that there is still an urgent need to identify VTE pathogenic genetic factors <sup>[9]</sup>.

With the development of gene chip technology, large-scale deep sequencing and bioinformatics analysis, scientists now have rich datasets for answering biological questions, including information on DEGs, pathways, and even targeted drugs associated with disease development <sup>[10]</sup>. Whole-exome sequencing and protein structure analysis can detect potentially important mutations that have not been reported, which is, hence, of vital significance for VTE patients to enrich anticoagulation and the use of catheter-directed thrombolytic therapy <sup>[11]</sup>.

In this study, we used the *GSE48000* and *GSE19151* datasets downloaded from the Gene Expression Omnibus to identify DEGs associated with VTE. GO and KEGG enrichment analyses of these DEGs were then performed. WES for a total of 25 VTE patients and 17 normal people was performed to screen out pathogenic mutations in VTE associated DEGs. Based on our findings, *HNMT*(chr2: 138759649 C>T, rs11558538) appeared to be a genetic susceptibility risk factor for VTE, representing a novel, potentially druggable target for the treatment of VTE.

## MATERIALS AND METHODS

### Patients samples

Twenty-five VTE and 17 normal patients admitted to participating hospitals from July 2015 to December 2018 were selected as study subjects. All patients were confirmed for VTE by B-ultrasound scan or CT examination, and basic characteristics of the recruited VTE patients were recorded, including gender, age, medication history, disease history, history of cardiovascular diseases, and history of chronic obstructive pulmonary disease. All patients signed informed consent documents, and this study was approved by the ethics committee of the Shanghai Institute of Planned Parenthood Research and ethical committees of participating hospitals.

### Extraction of datasets

Gene expression profiles were mined from the GEO database, which distributes high-throughput gene expression and other functional genomics datasets, using the following keywords: 'VTE' and 'Homo sapiens' <sup>[12]</sup>. Two datasets, *GSE48000* and *GSE19151*, were identified for this study. The gene expression profile dataset of *GSE48000* included 40 high-risk VTE cases and 25 healthy controls and had data on whole blood-derived RNA samples sequenced using a GPL10558 Illumina Human HT-12 V4.0 expression bead chip <sup>[13]</sup>. The *GSE19151* dataset was generated on the GPL571 (HG-U133A\_2) Affymetrix Human Genome U133A 2.0 Array platform. This dataset contained 70 adults with VTE cases and 63 healthy controls <sup>[14]</sup>.

### Preprocessing and repeatability tests of datasets

The original raw expression data at the probe level was downloaded as CEL files and pre-processed and normalized with RMA using the 'affy' package in R version 4.0.2, followed by converting the data into corresponding gene expression data based on the different platform specifications <sup>[15]</sup>. The Pearson's correlation coefficient was determined to validate the intra-group data repeatability and heatmap generation was visualized based on the 'heatmap' R package <sup>[16]</sup>. Principal Component Analysis (PCA) was conducted to view the clustering trends according to sample-to-sample distances using the 'ggord' package in R <sup>[17]</sup>.

### Identification of DEGs

The 'limma' package in R program was applied to screen DEGs between VTE samples and normal samples <sup>[18]</sup>. A two-tailed t-test was performed to examine DEGs by  $\log_2$  (Fold Change) >1 or <-1 and adjusted P value <0.05. Genes satisfying these conditions were grouped separately as DEGs by volcano plot in R <sup>[19]</sup>.

### Gene ontology and KEGG pathway enrichment analysis

GO is used to describe the Biological Processes (BPs), Molecular Functions (MFs), and Cellular Components (CCs) of gene products in a hierarchical ontology <sup>[20]</sup>. Signaling pathway analysis was conducted to map DEGS to the Kyoto Encyclopedia of Gene and Genomes (KEGG), which is a pathway-related database for systematic and comprehensive analysis of gene functions <sup>[21]</sup>. GO and KEGG pathway enrichment analyses were performed using the 'cluster Profiler' package in R version 4.0.2 and P values less than 0.05 were considered statistically significant <sup>[22]</sup>. A GO network was visualized using the Metascape database to validate our results <sup>[23]</sup>.

### Whole-exome sequencing for DEGs

DNA was extracted from each patient using a DNA extraction kit (Qiagen, Hilden, Germany) from whole blood. Library construction, WES, and data analysis were conducted by iGeneTech in Shanghai. Then, 200 ng of genomic DNA from each individual was sheared by Biorupter (Diagenode, Belgium) to acquire 150-200 bp fragments. The ends of DNA fragment were repaired and Illumina adapters were added (Fast Library Prep Kit, iGeneTech, Beijing, China).

After sequencing libraries were constructed, whole exomes were captured using the AIExome Enrichment Kit V1 (iGeneTech, Beijing, China) and libraries sequenced on an Illumina NovaSeq 6000 (Illumina, San Diego, CA) Next-Generation sequencing

platform in the 150 bp PE mode. Bioinformatics analysis was performed to analyze nonsynonymous mutations including Single-Nucleotide Polymorphisms (SNPs) and Insertion-Deletions (INDELS) using GATK (Genome Analysis Toolkit). All allele frequency data for DEG mutations were compared with the 1000 Genomes Project and Exome Aggregation Consortium ExAC.

### Protein structure modeling and molecular docking

The three-dimensional structure diagram of *HNMT* was generated using swiss model and pymol v2.4. The DGIdb database was used to select drugs based on the genes that served as promising targets [24].

### Molecular docking

Ligand docking of *HNMT* and drugs was performed with default parameters using AutoDock molecular docking software (version 4.2) and the coordinates and box size were finalized according to ligand location [25].

### Statistical analysis

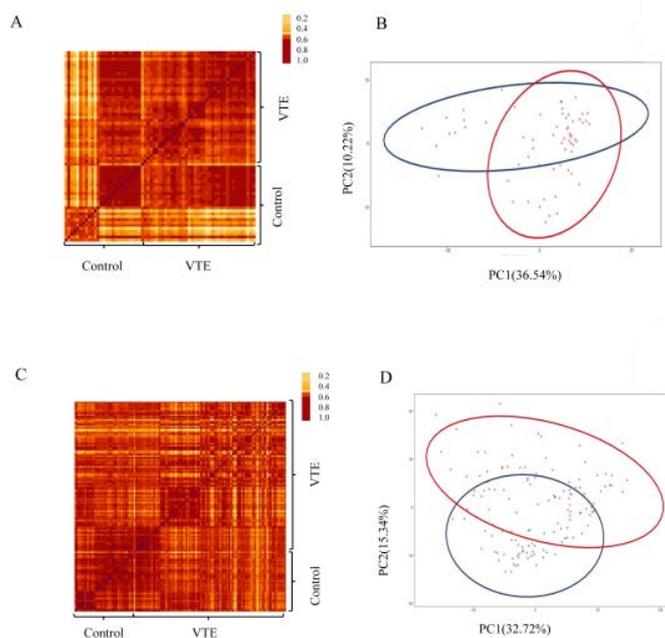
DEGs were selected based on t test. The whole genome GO categories and pathogenic mutations for these DEGs were identified using Fisher's exact test. The significance level for statistical tests was set at less than 0.05 ( $P < 0.05$ ).

## RESULTS

### Pearson's correlation testing and PCA

Pearson's correlation test showed strong correlations between VTE and control samples in the *GSE48000* dataset (Figure 1A). The PCA profile for the *GSE48000* data revealed that the distances between samples were small in the VTE groups and control groups, respectively (Figure 1B). Pearson's correlation analysis also indicated strong correlations for the *GSE19151* data among the samples in the VTE group and control group, respectively (Figure 1C). The close distance in the dimension of PCA illustrated the acceptable data repeatability between samples in the VTE group and control group for the *GSE19151* dataset (Figure 1D).

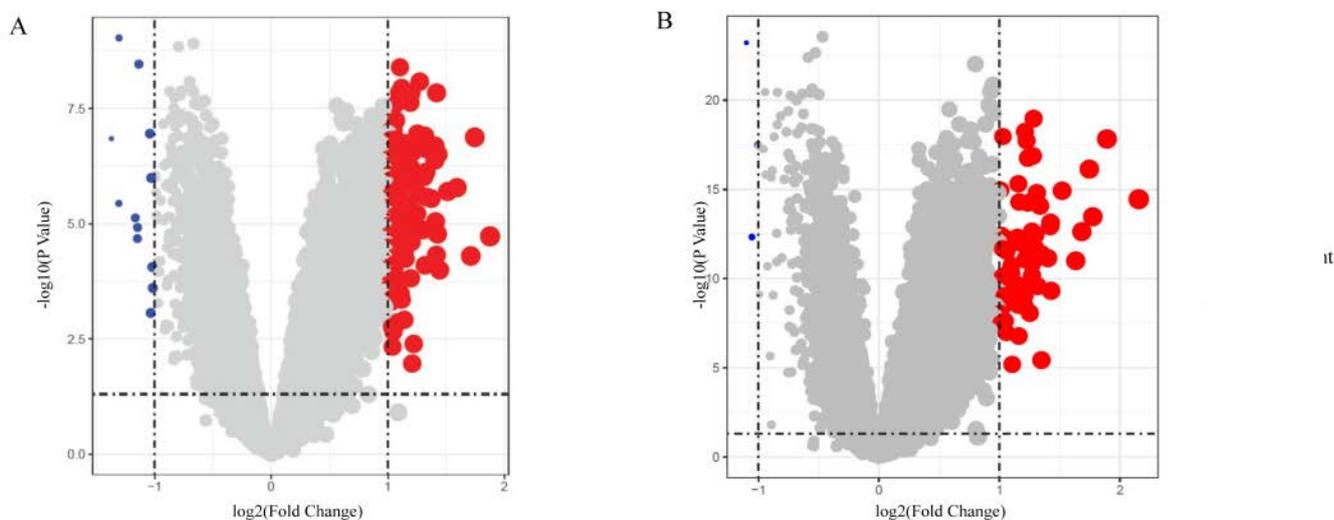
**Figure 1.** Pearson's correlation test and PCA on *GSE48000* and *GSE19151* data. (A) Pearson's correlation test for *GSE48000*. The color reflects the intensity of the correlation. (B) PCA of samples from the *GSE48000* dataset. Principal component 1 values for VTE samples are plotted on the X-axis, and principal component 2 values for control samples are plotted on the Y axis. The closer the distance between the two groups, the smaller the differences between the two groups. (C) Pearson's correlation test for *GSE19151*. The color reflects the intensity of the correlation. (D) PCA of samples from the *GSE19151* dataset. **Note:** Groups (■) Control; (■) VTE.



### Identification of DEGs in VTE

As shown in Figure 2, a total of 232 genes were designated as DEGs in the VTE group when compared with the control group. The volcano plots in this figure present the DEGs with a cutoff criteria of having an adjusted P-value  $< 0.05$  and  $|\log_2 \text{fold change}| > 1$  in the *GSE48000* and *GSE19151* datasets Figure 2. As examples of these differences, the top 10 differentially expressed genes are reported in Table 1.

**Figure 2.** Volcano plot of VTE DEGs. Red, upregulated DEGs with log2FC>1 and adjusted P-value<0.05. Green, downregulated DEGs with log2FC<-1 and adjusted P-value<0.05. (A) Volcano plot illustrating the DEGs of the GSE48000 dataset. (B) Volcano plot illustrating the DEGs of the GSE19151 dataset. **Note:** Regulated (●) Down; (●) No significant; (●) Up.



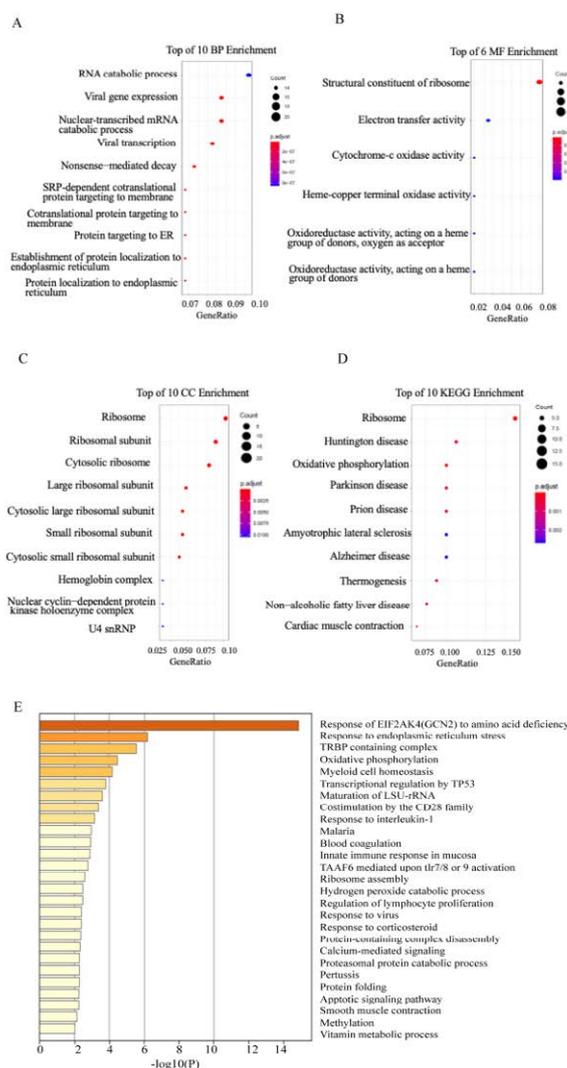
**Table 1.** Top 10 most downregulated DEGs in VTE.

Biochemical Tests	Biochemical Tests	Biochemical Tests	Biochemical Tests
<i>EVI2A</i>	2.1559596	$2.46 \times 10^{-13}$	Upregulation
<i>RPL9</i>	1.89195402	$1.48 \times 10^{-18}$	Upregulation
<i>IFI27</i>	1.8761832	$1.88 \times 10^{-5}$	Upregulation
<i>RPL31</i>	1.77436415	$3.44 \times 10^{-14}$	Upregulation
<i>NDUFA4</i>	1.74588189	$7.28 \times 10^{-17}$	Upregulation
<i>IGFBP1</i>	1.74512022	$1.33 \times 10^{-7}$	Upregulation
<i>SNORD8</i>	1.71070142	$5.02 \times 10^{-5}$	Upregulation
<i>RPS7</i>	1.68354151	$2.35 \times 10^{-13}$	Upregulation
<i>XK</i>	1.63316845	$1.00 \times 10^{-11}$	Upregulation
<i>RPS15A</i>	1.51993276	$1.20 \times 10^{-15}$	Upregulation
<i>FOS</i>	-1.37065523	$1.43 \times 10^{-7}$	Downregulation
<i>JMJD1C</i>	-1.30709903	$3.65 \times 10^{-6}$	Downregulation
<i>ZFP36L2</i>	-1.30546256	$9.36 \times 10^{-10}$	Downregulation
<i>CD46</i>	-1.16530934	$7.51 \times 10^{-6}$	Downregulation
<i>SNX10</i>	-1.14759082	$2.10 \times 10^{-5}$	Downregulation
<i>UBXN4</i>	-1.14648762	$1.22 \times 10^{-5}$	Downregulation
<i>DICER1</i>	-1.13523701	$3.46 \times 10^{-9}$	Downregulation
<i>LSP1</i>	-1.10000543	$5.98 \times 10^{-24}$	Downregulation
<i>TMEM259</i>	-1.05424088	$4.66 \times 10^{-13}$	Downregulation
<i>DCK</i>	-1.04088634	$1.11 \times 10^{-7}$	Downregulation

### Enrichment of DEGs by GO and KEGG analysis

Gene functional enrichment analysis was performed to analyze the biological connections of these DEGs. Results of Gene Ontology (GO) enrichment analysis revealed that RNA catabolic process, viral gene expression, SRP-dependent cotranslational protein targeting to membrane, and viral transcription were the main Biological Processes (BPs) and structural constituent of ribosome, cytochrome-c oxidase activity, and heme-copper terminal oxidase activity were the most enriched categories of molecular functions for these DEGs. The variations in Cell Component (CC) of DEGs were enriched largely in ribosome and hemoglobin complex. KEGG pathway analysis indicated that these DEGs were mainly involved in particular pathways, such as the ribosome, Huntington disease, and oxidative phosphorylation. Metascape was used to visualize these gene enrichment analyses to verify our results from R. We found that these DEGs were enriched in amino acid deficiency, ribosomal complex, oxidative phosphorylation, rRNA transcript, and blood coagulation Figure 3.

**Figure 3.** Bubble map for GO and KEGG pathway analyses of DEGs. P-value<0.05 was considered statistically significant. (A) Biological processes, (B) Molecular function, (C) Cell component, (D) KEGG pathways, (E) Heatmap of enriched terms across DEGs, colored by P-values, via Metascape.



### Identification of probable disease-causing DEGs by WES

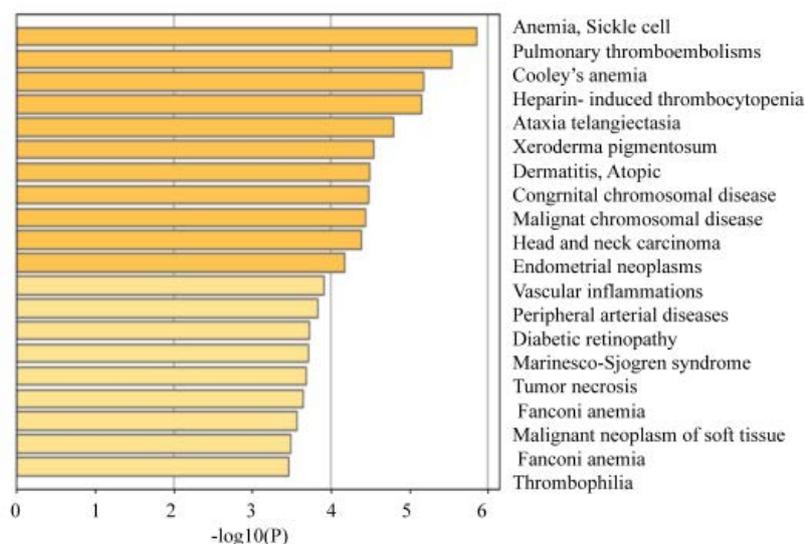
WES revealed 48 mutations of DEGs in the VTE group. The mutation types and the log2fold change are shown in Table 2. Because nonsynonymous mutations are most likely to affect protein function, we focused on the four SNP variants corresponding with four DEGs in the VTE group. These were *HNMT* (chr2: 138759649 C>T, rs11558538, adjusted P-value= $1.2 \times 10^{-9}$ ), *POLL* (chr10: 103340056 G>A, rs3730477, adjusted P-value= $5.12 \times 10^{-4}$ ), *ZNF292* (chr6: 87925827 A>G, rs9362415, adjusted P-value= $2.95 \times 10^{-8}$ ), and *DPCD* (chr10: 103361088 C>T, rs7874, adjusted P-value= $4.36 \times 10^{-5}$ ). The adjusted P-value of *HNMT* was the lowest in this study. Functional analysis showed that most disease-causing DEGs were involved in anemia, sickle cell, pulmonary thromboembolisms, heparin-induced thrombocytopenia, thrombophilia, and so on, as shown by Metascape functional analysis Figure 4. As summarized in Table 3, we found that *HNMT* was expressed in heparin-induced thrombocytopenia, dermatitis, and atopic cases, conditions that may have strong impacts on VTE.

**Table 2.** Probable disease-causing DEGs of VTE.

Gene	Chr	SNP	Mut_type	Mutation	Location	Position	Func.ref	Exonicfunc.ref	1 KG	ExAC_ALL	Cases(n=25)	Controls(n=17)	logFC	Adj.P.Val
<i>HNMT</i>	chr2	rs11558538	SNP	C/T	ontarget	138759649	exonic	nonsynonymous SNV	0.1	0.1017	1	5	1.05188	1.2×10 <sup>-9</sup>
<i>USP14</i>	chr18	rs56806027	SNP	T/A	flank150	204815	intronic		0.7		2	7	1.26505	0.000148
		rs57035428	InDel	T/TAAAAA	flank150	204816	intronic				2	7		
<i>SERPING1</i>	chr18	rs17072114	SNP	T/C	flank150	61584817	intronic		0.2		5	9	1.09247	0.000581
<i>UBXN4</i>	chr2	rs80198954	SNP	C/A	ontarget	136511842	exonic	synonymous SNV	0	0.019	6	12	-1.14649	0.000369
		rs74265494	SNP	G/A	ontarget	136511886	intronic		0	0.0191	6	12		
		rs372143998	InDel	GT/G	flank150	136527319	intronic			0.2468	6	11		
		rs200613240	InDel	T/TA	flank150	136529897	intronic				0	9		
		rs78878675	SNP	G/A	ontarget	136530157	intronic		0	0.019	6	12		
		rs78339162	SNP	A/G	flank150	136533993	intronic		0		6	11		
<i>ZNF271P</i>	chr18	rs12965288	SNP	C/A	ontarget	32888090	ncRNA_exonic		0.2		2	7	1.09133	1.09133
		rs34841246	SNP	C/A	ontarget	32888546	ncRNA_exonic		0.2		3	8		
<i>SELP</i>	chr1	rs35706397	InDel	T/TA	ontarget	169560727	intronic		0.3	0.3319	8	21	1.0297	0.000283
<i>GYPA</i>	chr4	rs62334651	SNP	T/C	flank150	145040784	intronic		0.4		12	2	1.04229	1.19×10 <sup>-08</sup>
		rs62334653	SNP	G/A	flank151	145041036	intronic				11	3		
<i>TMEM259</i>	chr19	rs2240161	SNP	A/G	ontarget	1011823	intronic		0.7	0.7045	14	15	-1.05424	1.52×10 <sup>-11</sup>
		rs7146	SNP	A/G	ontarget	1014398	exonic	synonymous SNV	0.7	0.6579	14	15		
<i>POLL</i>	chr10	rs1055364	SNP	C/A	ontarget	103338730	UTR3		0.2		0	4	1.14185	0.000512
		rs1055362	SNP	A/G	ontarget	103338733	UTR3		0.2		0	4		
		rs3730477	SNP	G/A	ontarget	103340056	exonic	nonsynonymous SNV	0.1	0.1678	0	4		
		rs3730476	SNP	A/G	ontarget	103340144	exonic	synonymous SNV	0.2	0.2306	0	4		
		rs3730475	SNP	A/G	ontarget	103340179	intronic		0.2	0.232	0	4		
		rs3730474	SNP	T/C	flank150	103340235	intronic		0.2		0	4		
		rs3730465	SNP	A/G	flank150	103343533	intronic		0.2		0	4		

		rs3730462	InDel	CTGTTG/C	ontarget	103345941	intronic		0.2	0.2281	0	4		
<i>ASTN1</i>	chr1	rs868002876	InDel	AT/A	ontarget	176913216	intronic		0.4		1	5	1.06975	0.001028
<i>UGGT1</i>	chr2	rs35069237	InDel	GT/G	ontarget	128949841	UTR3		0.3		7	12	1.14324	0.000712
<i>NFATC1</i>	chr18	rs8096658	SNP	C/G	flank150	77156537	intronic		0.3		0	4	1.09687	4.59E-05
		rs56376587	SNP	A/C	flank150	77160235	intronic		0.3		3	8		
<i>MTHFR</i>	chr1	rs11121832	SNP	T/C	flank150	11860120	intronic		0.8		18	17	1.10584	0.000173
<i>FCGR1B</i>	chr1	rs827371	SNP	T/C	flank150	120935661	intronic				4	8	1.0236	2.69×10 <sup>-11</sup>
<i>MAP3K8</i>	chr10	rs3034	SNP	G/A	flank150	30749895	UTR3		0.9		21	9	-1.01569	0.00025
<i>MGMT</i>	chr10	rs2782888	SNP	T/G	flank150	131265328	upstream		1		18	17	1.03576	0.000187
		rs55973415	SNP	G/A	flank150	131557750	intronic		0.2		6	10		
<i>ZNF2929</i>	chr6	rs563101504	InDel	GACACAC/G	ontarget	87925827	intronic				1	5	1.01862	2.95×10 <sup>-08</sup>
		rs9362415	SNP	A/G	ontarget	87968565	exonic	nonsynonymous SNV	0.6	0.5996	24	12		
		rs3734187	SNP	C/T	ontarget	87969737	exonic	synonymous SNV	0.6	0.5876	24	12		
		rs3812132	SNP	C/G	ontarget	87969737	exonic	synonymous SNV	0.7	0.5867	24	12		
		rs35541349	InDel	G/GA	flank150	87969737	UTR3				12	1		
<i>FOS</i>	chr14	rs1063169	SNP	G/T	flank150	75747118	intronic		0.1		8	0	-1.37066	3.96E-05
<i>ZNF346</i>	chr5	rs11448853	InDel	A/AG	flank150	176471286	intronic		0.5		11	14	1.0575	0.000198
<i>ERF</i>	chr19	rs61735151	SNP	G/A	ontarget	42753283	exonic	synonymous SNV	0.1	0.0878	25	13	1.03103	0.000499
<i>ALKBH89</i>	chr11	rs589316	SNP	G/A	flank150	107402887	intronic		0.4		3	9	1.0342	0.003568
		rs71488261	SNP	T/A	flank150	107422440	intronic		0.3		0	4		
<i>WDR55</i>	chr5	rs2251860	SNP	T/C	ontarget	140048209	exonic	synonymous SNV	0.5	0.4722	16	16	1.11147	0.000361
<i>AHSP</i>	chr16	rs10843	SNP	T/C	ontarget	31540030	UTR3		0.1	0.1185	3	8	1.2937	4.95×10 <sup>-11</sup>
<i>DPCD</i>	chr10	rs7911520	SNP	A/G	flank150	103354554	intronic		0.2		0	4	1.40465	4.36E-05
		rs7874	SNP	C/T	ontarget	103361088	exonic	nonsynonymous SNV	0.1	0.168	0	4		

**Figure 4.** Heatmap of enriched terms across disease-causing DEGs, via Metascape.



**Table 3.** Functional enrichment analysis of disease-causing DEGs using Metascape.

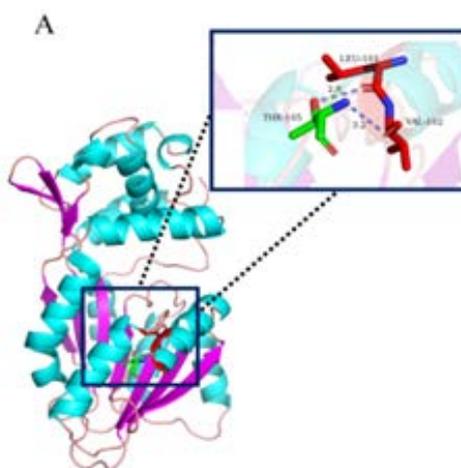
GO	Description	Log10p	Count	Genes
C0002895	Anemia Sickle cell	-5.9	6	FOS GYPA MTHFR SELP AHSP UGGT1
C0524702	Pulmonary thromboembolisms	-5.5	3	MTHFR SELP USP14
C0002875	Cooley's anemia	-5.2	4	GYPA MTHFR AHSP UGGT1
C0272285	Heparin-induced thrombocytopenia	-5.1	3	FCGR1B HNMT SELP
C0004135	Ataxia telangiectasia	-4.8	5	FOS GYPA MGMT MTHFR NFATC1
C0268138	Xeroderma pigmentosum	-4.5	3	MGMT MTHFR UGGT1
C0011615	Dermatitis, Atopic	-4.5	6	ASTN1 FOS HNMT MGMT MTHFR SELP
C0008626	Congenital-chromosomal disease	-4.5	6	FCGR1B FOS MGMT MTHFR NFATC1 SELP
C0278996	Malignant chromosomal disease	-4.4	6	FCGR1B FOS MGMT MTHFR NFATC1 SELP
C3887461	Head and neck carcinoma	-4.4	6	FCGR1B FOS MGMT MTHFR NFATC1 SELP
C0014170	Endometrial neoplasms	-4.2	4	MAP3K8 FOS MGMT MTHFR
C0947751	Vascular inflammations	-3.9	4	SERPING1 MAP3K8 FOS SELP
C1704436	Peripheral arterial diseases	-3.9	4	MAP3K8 FOS MTHFR SELP
C0011884	Diabetic retinopathy	-3.7	5	SERPING1 MAP3K8 FOS SELP MTHFR
C0024814	Marinesco-Sjogren syndrome	-3.7	3	MAP3K8 MGMT MTHFR
C0333516	Tumor necrosis	-3.7	4	FOS MGMT MTHFR SELP
C3469521	Fanconi anemia	-3.6	4	GYPA MGMT MTHFR SELP
C4551686	Malignant neoplasm of soft tissue	-3.6	5	MAP3K8 FOS MGMT MTHFR NFATC1
C0015625	Fanconi anemia	-3.5	4	GYPA MGMT MTHFR SELP
C0398623	Thrombophilia	-3.5	3	SERPING1 MTHFR SELP

**Protein structure and characterization of missense HNMT Mutations**

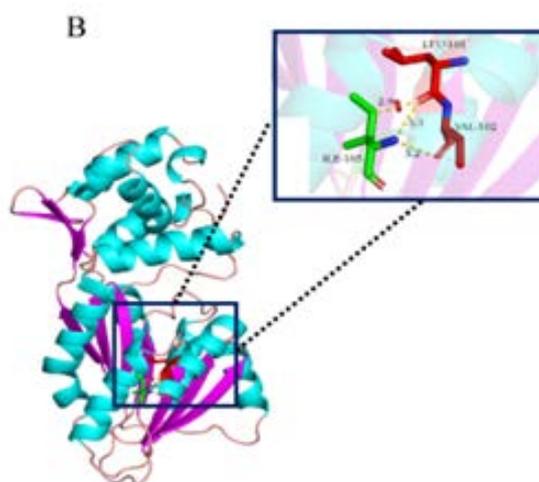
The Thr105Ile (rs11558538) polymorphism in the HNMT gene (ch2: 138759649 C>T, rs11558538, adjusted P-value=1.2 × 10<sup>-9</sup>) was the biggest difference identified in a gene, and should result in nonsense-mediated decay and loss function of this protein. The 3D location is shown in Figure 5A. The variant was positioned in the α-helix, where its side chain hydroxyl formed two hydrogen bonds with a backbone oxygen after mutation, causing a marked decrease in the levels of both HNMT enzymatic activity and immunoreactive protein [26, 27] Figure 5B. HNMT is an enzyme that has been implicated in neurotransmission by inactivating

histamine in the central nervous system [28]. However, histamine increases vascular permeability through the histamine H1 receptor to activate nerve endings, relaxing vascular smooth muscle [29].

**Figure 5A.** Diagram of the HNMT structure depicting the location of Thr105.



**Figure 5B.** Diagram of the HNMT structure depicting the location of the Ile105 mutation.



### Molecular docking

The drug–target interactions for *HNMT* were predicted using DGIdb, and the results are presented in Table 4, providing a theoretical therapeutic mechanism for VTE prevention. Six drugs targeting *HNMT* have been predicted for VTE, including Amodiaquine, Harmaline, Aspirin, Metoprine, Dabigatran, and Diphenhydramine. Molecular docking analysis was attempted to assess the potential noncovalent binding of *HNMT* with these small molecules drugs. In general, a lower binding energy indicated a stronger binding between *HNMT* and a compound.

Table 4 shows the six drugs that best interfaced with *HNMT*. To visualize these docking results, the 3D interaction diagrams of *HNMT* and their corresponding best-matched drugs were drawn, as shown in Figure 6.

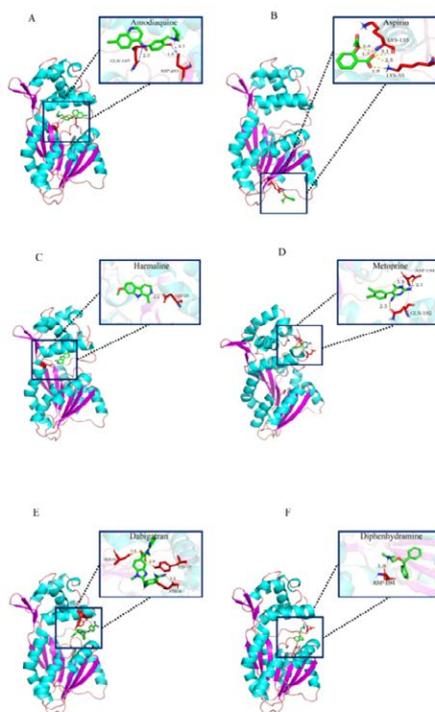
These drugs, such as Aspirin and Dabigatran, have been utilized to recanalize vessels and prevent thrombi growth clinically in VTE patients [30-32]. The 3D interaction diagram of Aspirin at the active site of *HNMT* revealed that this interaction was stable through forming hydrogen bonds with the key residues Lys55 and Lys135. Aspirin is commonly administered to inhibit platelet aggregation and prevent thrombus formation [33]. Additionally, three hydrogen bonds formed with residues Phe9, Tyr15, and Ser91 contributed to stabilizing the interaction between Dabigatran and *HNMT*. Dabigatran has been approved for use in orthopedic surgery, venous thromboprophylaxis, acute VTE treatment, and extended prevention of recurrent VTE [34]. Our data had shown that *HNMT* can potentially become a new target for VTE treatment. The current study was designed to investigate potential DEGs and genetic

variant of DEGs in VTE Figure 7.

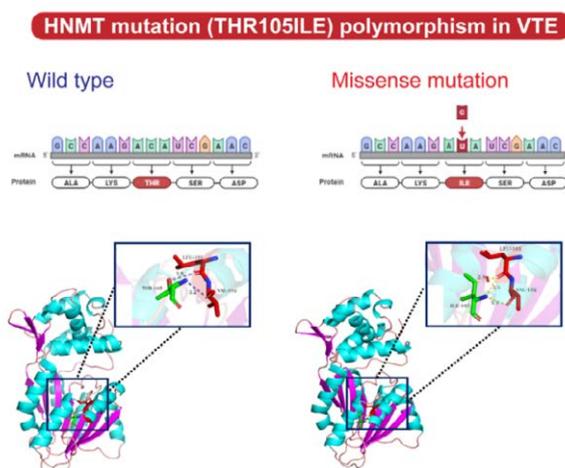
**Table 4.** Candidate drugs targeted HNMT.

Gene	Drug	Sources	PMIDs	Binding Energy (kcal. mol <sup>-1</sup> )	Binding Residues
<i>HNMT</i>	Amodiaquine	DrugBank	6789797	-2.48	GLN197 ASP203
<i>HNMT</i>	Aspirin	DrugBank	19178400	-3.24	LYS55 LYS135
<i>HNMT</i>	Harmaline	PharmGKB	1530666	-5.07	GLU28
<i>HNMT</i>	Metoprine	TTD;DTC	10592235	-3.65	GLN192 ASP194
<i>HNMT</i>	Dabigatran	TTD	-	-4	PHE9;TYR15;SER91
<i>HNMT</i>	Diphenhydramine	DrugBank	23896426	-4.69	ASP194

**Figure 6.** The 3D structure diagram of the drugs and *HNMT* with the active sites.(A) Structures of the pocket of binding between *HNMT* and Amodiaquine. (B) Structures of the binding between *HNMT* and Aspirin. (C) Structures of the binding between *HNMT* and Harmaline. (D) Structures of the binding between *HNMT* and Metoprine. (E) Structures of the binding between *HNMT* and Dabigatran. (F) Structures of the binding bet



**Figure 7.** Summary diagram regarding the genetic variant of DEGs in VTE.



## DISCUSSION

In the present study, transcriptomics and proteomics technology were used to explore the potential pathways and pathogenic mutations of VTE occurrence. High-throughput pharmacology and molecular docking may allow for the investigation of novel biomarkers for detecting this complex diseases <sup>[35]</sup>.

We first studied VTE by downloading transcriptome-wide expression data from the GEO database and a total of 232 DEGs were identified. Results of GO analysis of the gene enrichment in these datasets showed that the VTE-associated DEGs were significantly enriched in RNA catabolic process. Wang, HX found previously that RPL9, RPL35, and RPS7 were hub genes in the PPI network of GSE13985, which was used to identify potential markers of atherosclerosis development <sup>[36]</sup>. Interestingly, a study from Mi, YH reported that the major risk factors for atherothrombotic disease were also significantly associated with VTE, which contributes to the explanation of why atherosclerosis is an independent risk factor for VTE <sup>[37]</sup>. KEGG pathway analysis revealed that the DEGs related to VTE were mainly enriched in the ribosome pathway. Recent evidence has suggested that the ribosome affects the translation of platelets, platelet aggregation, and resultant thrombus formation <sup>[38,39]</sup>. It may be reasonable for us to then hypothesize that ribosomal proteins might have crucial functions in VTE development. However, there is no direct evidence that RNA catabolic processes and the ribosome pathway are directly involved in VTE.

To verify the above results, Whole-Exome Sequencing (WES) was performed to detect the pathogenic mutations of DEGs. POLL encodes the novel DNA polymerase lambda and the mutation of POLL (rs3730477) encoded R438W Pol λ leading to genomic instability and mutagenesis in cells <sup>[40]</sup>. DPCD (rs7874) is named from an uncharacterized genomic region surrounding POLL. DPCD is a novel gene in primary ciliary dyskinesia and severe cases can induce pulmonary embolism <sup>[41]</sup>. ZNF292 (rs9362415) acts as a transcription factor and plays an important role in DNA recognition and apoptosis regulation <sup>[42]</sup>. However, little is known about the role of DNA related functions in VTE. Notably, we discovered that *HNMT* (rs11558538) polymorphism was the greatest differentially expressed factor in this study. As is well-known, *HNMT* is implicated in neurotransmission by inactivating histamine, and histamine has been argued to relax vascular smooth muscle. From a protein structure analysis, we found that the Thr105Ile mutation results in hydrogen bonds in the structure of *HNMT* being disrupted, resulting in loss-of-function mutations <sup>[43]</sup>. The 3D structure diagram of *HNMT* showed quite different in protein conformations between the Thr105 and Ile105 variants. Furthermore, a list of drugs targeting *HNMT* with potential therapeutic efficacy against VTE were selected, most notably Aspirin and Dabigatran. As a consequence, we inferred that the high expression of mutated *HNMT* acted on vascular smooth muscle and may further promote vasoconstriction and thrombosis through RNA catabolic process and the ribosome. However, the mechanisms of these genes and drugs in VTE are still unclear. In future work, we hope to verify our conclusions experimentally to elucidate the effects of Thr105Ile (rs11558538) in *HNMT* for VTE.

## CONCLUSION

RNA catabolic process and ribosomes pathway identified through integrated bioinformatic analysis of GSE48000 and GSE19151 datasets may play crucial roles in the development of VTE. Additionally, Thr105Ile (rs11558538) polymorphism in *HNMT* was identified as a risk factor for VTE in the mechanism of damage and dysfunctional to the vascular endothelial cell and vascular smooth muscle. In the future, more in-depth investigation about the mechanism of these candidate genes is warranted for VTE.

## ACKNOWLEDGEMENTS

We thank the many who participated in this VTE study and funded this work.

## AUTHORS' CONTRIBUTIONS

JD were responsible for acquisition of data, analysis and interpretation of data. YZ and YC critically revised the work. YC were responsible for the conception and design of the study and final approval of the version to be submitted. The manuscript was written by JD. All authors read and approved the final manuscript.

## FUNDING

This project was supported by grants from the National Natural Science Foundation of China (Award No:81472990), the Clinical Research Project of Shanghai Municipal Health Commission (Award No:201840095) and the Clinical Research Project of Shanghai Municipal Health Commission (Award No:20214Y0332).

## AVAILABILITY OF DATA AND MATERIALS

Not applicable.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE

All patients participating in this study were informed, signed informed consents and voluntarily participated, and this study was approved by the ethics committee of the Shanghai Institute of Planned Parenthood Research and ethical committees of

participating hospitals.

### CONSENT FOR PUBLICATION

Not applicable.

### COMPETING INTERESTS

All authors declare that they have no competing interests.

### REFERENCES

1. Wang W, et al. MiR-150 promotes angiogenesis and proliferation of endothelial progenitor cells in deep venous thrombosis by targeting SRCIN1. *Microvasc Res.* 2019;123:35-41.
2. Evangelista MS, et al. venous thromboembolism and route of delivery: Review of the literature. *Rev Bras Ginecol Obstet.* 2018;40:156-162.
3. Xu Y, et al. Ethnoracial variations in venous thrombosis: Implications for management, and a call to action. *J Thromb Haemost.* 2020;19:30-40.
4. Lee LH, et al. Incidence of venous thromboembolism in Asian populations: A systematic review. *Thromb Haemost.* 2017;117:2243-2260.
5. Riedl J. Venous Thromboembolism in Brain Tumors: Risk factors, molecular mechanisms, and clinical challenges. *Semin Thromb Hemost.* 2019;45:334-341.
6. Moser G, et al. Maternal platelets—friend or foe of the human placenta? *Int J Mol Sci.* 2019;20:5639.
7. Zöller B, et al. Genetic risk factors for venous thromboembolism. *Expert Rev Hematol.* 2020;13:971-981.
8. Kumar DR, et al. Virchow's contribution to the understanding of thrombosis and cellular biology. *Clin Med Res.* 2010;8:168-172.
9. Rouhi-Broujeni H, et al. Association of homozygous thrombophilia polymorphisms and venous thromboembolism in shahrekord, Iran. *Tanaffos.* 2016;15:218-224.
10. Lindström S, et al. A large-scale exome array analysis of venous thromboembolism. *Genet Epidemiol.* 2019;43:449-457.
11. Almohammed OA, et al. Trends of cancer-associated Venous Thromboembolism (VTE) in the United States (2005-2014). *Thromb Res.* 2019;182:110-5.
12. Clough E. The Gene expression omnibus database. *Methods Mol Biol.* 2016;1418:93-110.
13. Lewis DA, et al. Whole blood gene expression profiles distinguish clinical phenotypes of venous thromboembolism. *Thromb Res.* 2015;135:659-665.
14. Lewis DA, et al. Whole blood gene expression analyses in patients with single versus recurrent venous thromboembolism. *Thromb Res.* 2011;128:536-540.
15. Gautier L, et al. affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics.* 2004;20:307-315.
16. Lun ATL, et al. Infrastructure for genomic interactions: Bioconductor classes for Hi-C, ChIA-PET and related experiments. *F1000Res.* 2016;20:950.
17. Jolliffe IT. Principal component analysis: A review and recent developments. *Philos Trans A Math Phys Eng Sci.* 2016;374:20150202.
18. Ritchie ME, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:47.
19. Kumar N, et al. Robust volcano plot: identification of differential metabolites in the presence of outliers. *BMC Bioinformatics.* 2018;19:128.
20. Zhao Y, et al. Gene function prediction based on Gene Ontology Hierarchy Preserving Hashing. *Genomics.* 2019;111:334-342.
21. Kanehisa M, et al. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 2016;44:457-462.
22. Yu G, et al. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS.* 2012;16:284-287.
23. Zhou Y, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* 2019;10:1523.

24. Cotto KC, et al. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.* 2018;46:1068-1073.
25. Morris GM, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem.* 2009;30:2785-2791.
26. Chen Y, et al. Determining the Effect of the *HNMT*, *STK39*, and *NMD3* polymorphisms on the incidence of parkinson's disease, amyotrophic lateral sclerosis, and multiple system atrophy in chinese populations. *J Mol Neurosci.* 2018;64:574-580.
27. Jiménez-Jiménez FJ, et al. Thr105Ile (rs11558538) polymorphism in the histamine-1-methyl-transferase (*HNMT*) gene and risk for restless legs syndrome. *J Neural Transm (Vienna).* 2017;124:285-291.
28. Meza-Velázquez R, et al. Association of diamine oxidase and histamine N-methyltransferase polymorphisms with presence of migraine in a group of Mexican mothers of children with allergies. *Neurología.* 2017;32:500-507.
29. Huang M, et al. Critical roles of TRPV2 channels, histamine H1 and adenosine A1 receptors in the initiation of acupoint signals for acupuncture analgesia. *Sci Rep.* 2018;8:6523.
30. Hanafy AS, et al. Randomized controlled trial of rivaroxaban versus warfarin in the management of acute non-neoplastic portal vein thrombosis. *Vascul Pharmacol.* 2019;113:86-91.
31. Tan ZK, et al. PO-10-Lung cancer with tumour thrombosis: A poor prognosticator. *Thromb Res.* 2016;140:180.
32. Suzuki K, et al. Bevacizumab-induced aortic arterial thrombosis. *Intern Med.* 2018;57:2987-2990.
33. Kang G. Aspirin for postoperative venous thromboembolism prophylaxis. *J Gen Intern Med.* 2020;35:2778.
34. Schulman S, et al. Dabigatran versus warfarin in the treatment of acute venous thromboembolism. *N Engl J Med.* 2009;361:2342-2352.
35. Azad RK. Metabolomics technology and bioinformatics for precision medicine. *Briefings in bioinformatics.* 2019;20:1957-1971.
36. Wang HX. Prediction of genetic risk factors of atherosclerosis using various bioinformatic tools. *Genet Mol Res.* 2016;15:2.
37. Mi Y, et al. Venous thromboembolism has the same risk factors as atherosclerosis: A PRISMA-compliant systemic review and meta-analysis. *Medicine (Baltimore).* 2016;95:4495.
38. Rowley JW. Ribosomes in platelets protect the messenger. *Blood.* 2017;129:2343-2345.
39. Wang H, et al. The microvesicle/CD36 complex triggers a prothrombotic phenotype in patients with non-valvular atrial fibrillation. *J Cell Mol Med.* 2020;24:7331-7340.
40. Nemec AA, et al. Estrogen drives cellular transformation and mutagenesis in cells expressing the breast cancer-associated R438W DNA polymerase lambda protein. *Mol Cancer Res.* 2016;14:1068-1077.
41. Nandhakumar A. Acute hypoxemia in a parturient with primary ciliary dyskinesia following the administration of intravenous oxytocin: a case report. *Can J Anaesth.* 2013;60:1218-1221.
42. Mirzaa GM, et al. De novo and inherited variants in ZNF292 underlie a neurodevelopmental disorder with features of autism spectrum disorder. *Genet Med.* 2020;22:538-546.
43. Verhoeven WMA, et al. Adult male patient with severe intellectual disability caused by a homozygous mutation in the *HNMT* gene. *BMJ Case Rep.* 2020;13:235972.